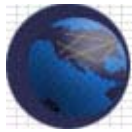




Integrative Cancer Research Special Interest Group Teleconference

Microarray Repository SIG Meeting Minutes

Date, Time & Location:	May 5, 2004 2:00 – 3:00 EDT
Attendees:	Patrick McConnell – Duke (funded developer) Tom Moloshok – Fox Chase (funded developer) Steve Enkemann – Moffitt (funded adopter) Gary Bader – Sloan (funded developer, funded adopter) Elana for Judith Goldberg – New York (funded adopter) Ajay Jain – UC San Francisco (funded developer) Taku Tokuyasu – UC San Francisco (funded developer) Suzanne Conzen – U of C (funded developer) Xinmin Li – U of C (funded developer) Terry Braun – Holden (funded developer) Erin Brothers – Holden (funded developer) David Fenstermacher – Penn (funded adopter) Louise Showe – Wistar (funded adopter) Michael Showe – Wistar (funded adopter) Scott Oster – Ohio State David Kane – SRA Mervi Heiskanen – NCICB Phan Winer – BAH Juli Klemm – BAH
Introduction:	<u>Roll-call, open meeting, review meeting goals</u> <ul style="list-style-type: none">- Establish goals and priorities for this SIG- For match making purposes, Developers and Adopters will be asked to give a brief statement of their capabilities and interests, respectively- Identify and define additional research
Overview Discussion:	<u>Review goals and objectives of Microarray Repository SIG</u> <p>Scope: Focus on tools and systems to manage microarray data.</p> <u>Open Discussion</u> <ul style="list-style-type: none">- There is the question of whether to store microarray data as BLOBS/flat files or whether to store the data relationally.<ul style="list-style-type: none">o Storing the data relationally allows powerful queries to be performed on the data - “Show me all experiments where this gene is upregulated for this phenotype.”o Normalization is an important issue to consider when performing queries across experiments.o Flat file data is required by most external analysis toolso May need two different databases – one for raw data and one for processed data- Many adopters have an urgent need for a general repository with the flexibility to store all relevant information, including gene data, sample data, and images.<ul style="list-style-type: none">o Organizing legacy data is a critical need



Integrative Cancer Research Special Interest Group Teleconference

- Replace the “shoeboxes of CD-ROMS”
- Comparing gene expression data across technologies should be considered: SAGE, QT-PCR, EST profiling, etc.
- MIAME standards
 - People like MIAME to a point – they like it when someone else populates the information but many users find it burdensome to enter this data.
 - Should make it easy for users to enter this data and find ways to incent them – e.g. require the information up front at the time of sample submission to the microarray facility
 - Standards such as MIAME are good, but they can get in the way of getting work done. Important to have some flexibility
 - MIAME standards aren’t in and of themselves burdensome if you “work as you go” but most people don’t pay attention to them until that have to submit their data.
 - Keep in mind the use cases: Understand how the end users will use the data and create a tools flexible enough that new use cases can be added as needed.
- Data quality descriptors should be included with datasets. A database for a large number of contributors will have data sets of widely-ranging quality.
 - Let the user decide what data quality should be included in their query
 - Can consider an analogy with GenBank (varied quality) and RefSeq (highest quality)

High-level review of Center interests in this SIG

Developers

Iowa: CED is a tool integrate expression data with clinical data and to ask interesting relational questions. Three-pronged approach: clinical, expression, integration. Technologies include Affymetrix, glass, EST sequencing, SAGE

CCR: MADB currently has 5 – 6 instances, ~1000 users. It supports the processing of raw data with a suite of algorithms and filters. Can recreate results from original data. Supports both glass and Affymetrix. Currently adding support for MIAME, but users will not be required to submit this data.

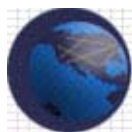
NCICB: caArray. First release will be mid-September for the database; tools will be separate. There will be a central repository at NCI but users can also create local instances. Supports controlled vocabularies and the MIAME standard – MIAME is not enforced. Uses the caCORE infrastructure to integrate other data types including gene annotations. NCI will be running this on Oracle but any database can be used. Is written in Java so can be compiled to run on anything.

Adopters

Duke: Acute need for a microarray repository

Fox Chase: Currently using a commercial system – BioDiscovery – but it has been slow to roll out. Would be interested in an alternative.

Sloan: Attending this SIG to see what is available. Currently have a basic data storage system for their Affymetrix data, some metadata. Require the capability to



Integrative Cancer Research Special Interest Group Teleconference

release data publicly, to easily access the data and for some basic analysis functions.

New York: In need of a repository for their Affymetrix and 2-color data. Have the need to share, access, and manipulate the data.

UC San Francisco: As a "developer adopter", looking to interface their tools with a standard microarray repository, likely caArray. As an "end user adopter" looking for a system to store many gene expression data types: Affymetrix, 2-color spotted, 3-color CGH and custom oligo-based arrays.

U of C: Possibly looking for a replacement for their MADAM repository that is no longer under development. Using cDNA glass slides as well as Affymetrix technology. Have good experience with repository development and would like to become a developer for caBIG.

Penn: Technologies include Affymetrix, oligo arrays, custom arrays, CGH, and Affymetrix SNP chips. Interested in storing the data relationally with a web-based system to access the data. Would like to associate with clinical data.

Wistar: Currently storing Affymetrix data in Excel spreadsheets – in need of a repository. Would like to store the images for possible re-analysis. All clinical data is stored separately.

Moffitt: Have an EBI database in house; has been difficult and time-consuming to load the data. Original goal was to find a tool to load and capture the relevant data to this database, but would be willing to start from scratch, if necessary. Have mostly Affymetrix data, some spotted array data.

CCR: Have data from a large number of technologies including Q-PCR, SKY, high-density protein arrays, methylation studies. Need analysis tools to do an integrated analysis of this data.

Other discussion items:

caArray web demo

- Many in the group said they would benefit from an on-line presentation of caArray. The proposed date of May 19 was bad for many people. Juli will send out alternative dates and choose the one that fits best for the most participants.
- Participants should inform Juli if there are other systems that might be demoed to the group.

Mission Statement

Steve Enkemann from Moffitt volunteered to draft a Mission Statement for this SIG. This draft will be circulated to the group for comments.

Regular meeting schedule

The group will meet on the first Wednesday of each month at 2:00 Eastern. Ongoing frequency will be revisited if necessary.

Notes and additional information

Meeting minutes will be emailed and posted on the caBIG on-line forum:

http://ncicbforums.nci.nih.gov/forums/cabigforum/lfs/icrlfs/SIGs/microarray_rep.

Developers should also post additional information about their tools here

Action Items:

Name Responsible	Action Item	Date Due	Notes
Steve Enkemann	Draft Microarray SIG mission statement	5/21/04 (approx)	



Integrative Cancer Research Special Interest Group Teleconference

	Juli Klemm	Distribute meeting minutes	5/7/04	
	Juli Klemm	Propose possible dates for caArray on-line demo	5/10/04	
	Juli Klemm	Schedule ongoing meetings	5/14/04	
	All developers	Post relevant additional information to the caBIG on-line forum (see link above)	5/14/04	